(a) Consider the usual linear model, where $\mathbf{Y = X\beta + \varepsilon}$. We now compare two regressions, which differ in how many variables are included in the matrix $\mathbf{X}$. In the full (unrestricted) model $\mathbf{p_1}$ regressors are included. In the restricted model only a subset of $\mathbf{p_0 < p_1}$ regressors are included. Show that the smallest model is preferred according to the **AIC** if

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p_1 - p_0)}$$

Solution

The Akaike information criterion (**AIC**) are defined as follows, where p is the number of included regressors and s^2 is the maximum likelihood estimator of the error variance in the model with p regressors:

$$\mathbf{AIC} = \log(s^2) + \frac{2p}{n},$$

Thus for full (unrestricted) and restricted models the **AIC** equal to $\mathbf{AIC_1} = \log(s_1^2) + \frac{2p_1}{n}$ [1] and $\mathbf{AIC_0} = \log(s_0^2) + \frac{2p_0}{n}$ respectively**.** We find s_1^2 and s_0^2 from **AIC₁** and **AIC₀**:

$\mathbf{s_1^2} = \mathbf{e}^{\mathrm{AIC}_1 - \frac{2p_1}{n}}$ and $\mathbf{s_0^2} = \mathbf{e}^{\mathrm{AIC}_0 - \frac{2p_0}{n}}$**.** Expressions s_1^2 and s_0^2 are used in the inequality **(a):**

$$\frac{\mathbf{e}^{\mathrm{AIC}_0 - \frac{2p_0}{n}}}{\mathbf{e}^{\mathrm{AIC}_1 - \frac{2p_1}{n}}} < e^{\frac{2}{n}(p_1 - p_0)}$$

$$\mathbf{e}^{\mathrm{AIC}_0 - \frac{2p_0}{n} - (\mathrm{AIC}_1 - \frac{2p_1}{n})} < e^{\frac{2}{n}(p_1 - p_0)}$$

$$\mathbf{e}^{\mathrm{AIC}_0 - \frac{2p_0}{n} - \mathrm{AIC}_1 + \frac{2p_1}{n}} < e^{\frac{2}{n}(p_1 - p_0)}$$

$$\mathbf{e}^{\mathrm{AIC}_0 - \mathrm{AIC}_1 + \frac{2}{n}(p_1 - p_0)} < \mathbf{e}^{\frac{2}{n}(p_1 - p_0)}$$

The restricted model is preferred above the unrestricted model by AIC, in the sense that AIC₀ < AIC₁, if the F-test in is smaller than 2.

Therefore, considering condition **AIC₀ < AIC₁** inequality $\mathbf{e}^{\mathrm{AIC}_0 - \mathrm{AIC}_1 + \frac{2}{n}(p_1 - p_0)} < \mathbf{e}^{\frac{2}{n}(p_1 - p_0)}$ is performed.

---

[1] This is the natural logarithm i.e., **Ln.**

(b) Argue that for very large values of **n** the inequality of **(a)** is equal to the condition

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p_1 - p_0)$$

Use that $e^x \approx 1+x$ for small values of **x**.

Solution

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p_1 - p_0)$$

$$\frac{s_0^2}{s_1^2} - 1 < \frac{2}{n}(p_1 - p_0)$$

In this inequality $\frac{s_0^2}{s_1^2} < \frac{2}{n}(p_1 - p_0) + 1$ instead of the expression $\frac{s_0^2}{s_1^2}$ we use the expression $e^{\frac{2}{n}(p_1 - p_0)}$ from the inequality of **(a).**

$$e^{\frac{2}{n}(p_1 - p_0)} < \frac{2}{n}(p_1 - p_0) + 1$$

For very large values of **n** i.e., $n \to \infty$ the values of $\frac{2}{n} = 0.$

Therefore,

$$e^{0*(p_1 - p_0)} < 0 * (p_1 - p_0) + 1$$

$$e^0 < 1$$

$$1 < 1$$

It should be noted that the expression $\frac{s_0^2}{s_1^2}$ is less than the expression $e^{\frac{2}{n}(p_1 - p_0)}$ i.e., $\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p_1 - p_0)}$**.** Thus, $\frac{s_0^2}{s_1^2} < 1$ and inequality $\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p_1 - p_0)$ is performed.

(c) Show that for very large values of **n** the condition in **(b)** is approximately equal to

$$\frac{e'_R e_R - e'_U e_U}{e'_U e_U} < \frac{2}{n}(p_1 - p_0)$$

Solution

We are using the following formula:

$$s_1^2 = \frac{1}{n-r} e'_U e_U \ , s_0^2 = \frac{1}{n-(r-g)} e'_R e_R$$

$$e'_U e_U = (n-r)s_1^2, e'_R e_R = (n-(r-g))s_0^2$$

$$\frac{(n-(r-g))s_0^2 - (n-r)s_1^2}{(n-r)s_1^2} < \frac{2}{n}(p_1 - p_0)$$

$$\frac{(n-(r-g))s_0^2}{(n-r)s_1^2} - 1 < \frac{2}{n}(p_1 - p_0)$$

In this inequality $\frac{(n-(r-g))s_0^2}{(n-r)s_1^2} - 1 < \frac{2}{n}(p_1 - p_0)$ instead of the expression $\frac{2}{n}(p_1 - p_0)$ we use the expression $\frac{s_0^2}{s_1^2} - 1$ from the inequality of **(b).** It should be noted that the expression $\frac{s_0^2}{s_1^2} - 1$ is less than the expression $\frac{2}{n}(p_1 - p_0)$.

$$\frac{(n-(r-g))s_0^2}{(n-r)s_1^2} - 1 < \frac{s_0^2}{s_1^2} - 1$$

$$\frac{(n-(r-g))s_0^2}{(n-r)s_1^2} < \frac{s_0^2}{s_1^2}$$

Due to the fact that the $\big(n-(r-g)\big) < (n-r)$, therefore the inequality $\frac{(n-(r-g))s_0^2}{(n-r)s_1^2} < \frac{s_0^2}{s_1^2}$ is performed.

Considering above analysis the inequality $\frac{e'_R e_R - e'_U e_U}{e'_U e_U} < \frac{2}{n}(p_1 - p_0)$ is performed.

(d) Finally, show that the inequality from (c) is approximately equivalent to an F-test with critical value 2, for large sample sizes.

Solution

$$\boldsymbol{F = \frac{(e_R' e_R - e_U' e_U)/g}{e_U' e_U/(n-r)}}$$

$$\boldsymbol{F = \frac{e_R' e_R - e_U' e_U}{e_U' e_U} * \frac{1}{g(n-r)}}$$

Due to the fact that the $\frac{e_R' e_R - e_U' e_U}{e_U' e_U} < \frac{2}{n}(\boldsymbol{p_1 - p_0})$**,** we must to argue that inequality $\frac{e_R' e_R - e_U' e_U}{e_U' e_U} * \frac{1}{g(n-r)} < \frac{e_R' e_R - e_U' e_U}{e_U' e_U}$ is performed. If $\frac{1}{g(n-r)} < \boldsymbol{1}$**,** the inequality $\frac{e_R' e_R - e_U' e_U}{e_U' e_U} * \frac{1}{g(n-r)} < \frac{e_R' e_R - e_U' e_U}{e_U' e_U}$ is true.